Applied Probability Trust (16 October 2008)

OPTIMAL CO-ADAPTED COUPLING FOR THE SYMMETRIC RANDOM WALK ON THE HYPERCUBE

STEPHEN CONNOR,* *University of Warwick*

SAUL JACKA,* *University of Warwick*

Abstract

Let X and Y be two simple symmetric continuous-time random walks on the vertices of the n-dimensional hypercube, $\mathbb{Z}_2^n$. We consider the class of co-adapted couplings of these processes, and describe an intuitive coupling which is shown to be the fastest in this class.

Keywords: Optimal coupling; co-adapted; stochastic minimum; hypercube

2000 Mathematics Subject Classification: Primary 93E20

Secondary 60J27

1. Introduction

Let $\mathbb{Z}_2^n$ be the group of binary n-tuples under coordinate-wise addition modulo 2: this can be viewed as the set of vertices of an n-dimensional hypercube. For $x \in \mathbb{Z}_2^n$, we write $x = (x(1), \ldots, x(n))$, and define elements $\{e_i\}_0^n$ by

$$e_0 = (0, \ldots, 0); \quad e_i(k) = \mathbf{1}_{[i=k]},\ i = 1, \ldots, n,$$

where $\mathbf{1}$ denotes the indicator function. For $x, y \in \mathbb{Z}_2^n$ let

$$|x - y| = \sum_{i=1}^{n} |x(i) - y(i)|$$

denote the Hamming distance between x and y.

A continuous-time random walk X on $\mathbb{Z}_2^n$ may be defined using a marked Poisson process Λ of rate n, with marks distributed uniformly on the set $\{1, 2, \ldots, n\}$: the i^{th} coordinate of X is flipped to its opposite value (zero or one) at incident times of Λ for which the corresponding mark is equal to i. We write $\mathcal{L}(X_t)$ for the law of X at time t. The unique equilibrium distribution of X is the uniform distribution on $\mathbb{Z}_2^n$.

* Postal address: Department of Statistics, University of Warwick, Coventry, CV4 7AL. UK

Suppose that we now wish to couple two such random walks, X and Y, starting from different states.

Definition 1.1. A *coupling* of X and Y is a process (X', Y') on $\mathbb{Z}_2^n \times \mathbb{Z}_2^n$ such that

$$X' \stackrel{\mathcal{D}}{=} X \quad \text{and} \quad Y' \stackrel{\mathcal{D}}{=} Y \,.$$

That is, viewed marginally, X' behaves as a version of X, and Y' as a version of Y.

For any coupling strategy c, write (X_t^c, Y_t^c) for the value at t of the pair of processes X^c and Y^c driven by strategy c, although this superscript notation may be dropped when no confusion can arise. (We assume throughout that (X^c, Y^c) *is* a coupling of X and Y.) We then define the coupling time by

$$\tau^c = \inf \{ t \geq 0 \,:\, X_s^c = Y_s^c \;\; \forall \, s \geq t \} \,.$$

Note that in general this is not necessarily a stopping time for either of the marginal processes, nor even for the joint process. For $t \geq 0$, let

$$U_t^c = \{ 1 \leq i \leq n \,:\, X_t^c(i) \neq Y_t^c(i) \}$$

denote the set of unmatched coordinates at time t, and let

$$M_t^c = \{ 1 \leq i \leq n \,:\, X_t^c(i) = Y_t^c(i) \}$$

be its complement. A simple coupling technique appears in [1], and may be described as follows:

- if $X(i)$ flips at time t, with $i \in M_t$, then also flip coordinate $Y(i)$ at time t (matched coordinates are always made to move synchronously);
- if $|U_t| > 1$ and $X(i)$ flips at time t, with $i \in U_t$, also flip coordinate $Y(j)$ at time t, where j is chosen uniformly at random from the set $U_t \setminus \{i\}$;
- else, if $U_t = \{i\}$ contains only one element, allow coordinates $X(i)$ and $Y(i)$ to evolve independently of each other until this final match is made.

This defines a valid coupling of X and Y, for which existing coordinate matches are maintained and new matches made in pairs when $|U_t| \geq 2$. It is also an example of a *co-adapted* coupling.

Definition 1.2. A coupling (X^c, Y^c) is called *co-adapted* if there exists a filtration $(\mathcal{F}_t)_{t\geq 0}$ such that

1. X^c and Y^c are both adapted to $(\mathcal{F}_t)_{t\geq 0}$
2. for any $0 \leq s \leq t$,

$$\mathcal{L}\left(X^c_t \,|\, \mathcal{F}_s\right) = \mathcal{L}\left(X^c_t \,|\, X^c_s\right) \quad \text{and} \quad \mathcal{L}\left(Y^c_t \,|\, \mathcal{F}_s\right) = \mathcal{L}\left(Y^c_t \,|\, Y^c_s\right) \,.$$

In other words, (X^c, Y^c) is co-adapted if X^c and Y^c are both Markov with respect to a common filtration, $(\mathcal{F}_t)_{t\geq 0}$. Note that this definition does *not* imply that the joint process (X^c, Y^c) is Markovian, however. If (X^c, Y^c) is co-adapted then the coupling time is a randomised stopping time with respect to the individual chains, and it suffices to study the first *collision* time of the two chains (since it is then always possible to make X^c and Y^c agree from this time onwards).

In this paper we search for the best possible coupling of the random walks X and Y on $\mathbb{Z}_2^n$ within the class $\mathcal{C}$ of all co-adapted couplings.

2. Co-adapted couplings for random walks on $\mathbb{Z}_2^n$

In order to find the optimal co-adapted coupling of X and Y, it is first necessary to be able to describe a general coupling strategy $c \in \mathcal{C}$. To this end, let Λ_{ij} $(0 \leq i, j \leq n)$ be independent unit-rate marked Poisson processes, with marks W_{ij} chosen uniformly on the interval $[0, 1]$. We let $(\mathcal{F}_t)_{t\geq 0}$ be any filtration satisfying

$$\sigma\left\{\bigcup_{i,j}\Lambda_{ij}(s), \bigcup_{i,j}W_{ij}(s) \,:\, s \leq t\right\} \subseteq \mathcal{F}_t, \quad \forall\, t \geq 0\,.$$

The transitions of X^c and Y^c will be driven by the marked Poisson processes, and controlled by a process $\{Q^c(t)\}_{t\geq 0}$ which is adapted to $(\mathcal{F}_t)_{t\geq 0}$. Here, $Q^c(t) = \left\{q^c_{ij}(t) \,:\, 1 \leq i, j, \leq n\right\}$ is a $n \times n$ doubly sub-stochastic matrix. Such a matrix implicitly defines terms $\left\{q^c_{0j}(t) \,:\, 1 \leq j \leq n\right\}$ and $\{q^c_{i0}(t) \,:\, 1 \leq i \leq n\}$ such that

$$\sum_{i=0}^{n} q^c_{ij}(t) = 1 \quad \text{for all } 1 \leq j \leq n \text{ and } t \geq 0\,, \tag{2.1}$$

$$\text{and} \quad \sum_{j=0}^{n} q^c_{ij}(t) = 1 \quad \text{for all } 1 \leq i \leq n \text{ and } t \geq 0\,. \tag{2.2}$$

For convenience we also define $q^c_{00}(t) = 0$ for all $t \geq 0$.

Note that any co-adapted coupling (X^c, Y^c) must satisfy the following three constraints, all of which are due to the marginal processes $X^c(i)$ $(i = 1, \ldots, n)$ being independent unit rate Poisson processes (and similarly for the processes $Y^c(i)$):

1. At any instant the number of jumps by the process (X^c, Y^c) cannot exceed two (one on X^c and one on Y^c);

2. All single and double jumps must have rates bounded above by one;

3. For all $i = 1, \ldots, n$, the *total* rate at which $X^c(i)$ jumps must equal one.

A general co-adapted coupling for X and Y may therefore be defined as follows: if there is a jump in the process Λ_{ij} at time $t \geq 0$, *and* the mark $W_{ij}(t)$ satisfies $W_{ij}(t) \leq q_{ij}(t)$, then set $X^c_t = X^c_{t-} + e_i \pmod 2$ and $Y^c_t = Y^c_{t-} + e_j \pmod 2$. Note that if i (respectively j) equals zero, then $X^c_t = X^c_{t-}$ (respectively, $Y^c_t = Y^c_{t-}$), since $e_0 = (0, \ldots, 0)$.

From this construction it follows directly that X^c and Y^c both have the correct marginal transition rates to be continuous-time simple random walks on $\mathbb{Z}^n_2$ as described above, and are co-adapted.

3. Optimal coupling

Our proposed optimal coupling strategy, $\hat{c}$, is very simple to describe, and depends only upon the number of unmatched coordinates of X and Y. Let $N_t = |U_t|$ denote the value of this number at time t. Strategy $\hat{c}$ may be summarised as follows:

- matched coordinates are always made to move synchronously (thus $N^{\hat{c}}$ is a decreasing process);
- if N is odd, all unmatched coordinates of X and Y are made to evolve independently until N becomes even;
- if N is even, unmatched coordinates are coupled in pairs - when an unmatched coordinate on X flips (thereby making a new match), a different, uniformly chosen, unmatched coordinate on Y is forced to flip at the same instant (making a total of two new matches).

Note the similarity between $\hat{c}$ and the coupling of Aldous described in Section 1: if N is even these strategies are identical; if N is odd however, $\hat{c}$ seeks to restore the parity of N as fast as possible, whereas Aldous's coupling continues to couple unmatched coordinates in pairs until $N = 1$.

Definition 3.1. The matrix process $\hat{Q}$ corresponding to the coupling $\hat{c}$ is as follows:

- $\hat{q}_{ii}(t) = 1$ for all $i \in M_t$ and for all $t \geq 0$;
- if N_t is odd, $\hat{q}_{i0}(t) = \hat{q}_{0i}(t) = 1$ for all $i \in U_t$;
- if N_t is even, $\hat{q}_{i0}(t) = \hat{q}_{0i}(t) = \hat{q}_{ii}(t) = 0$ for all $i \in U_t$, and

$$\hat{q}_{ij} = \frac{1}{|U_t| - 1} \quad \text{for all distinct } i, j \in U_t \,.$$

The coupling time under $\hat{c}$, when $(X_0, Y_0) = (x, y)$, can thus be expressed as follows:

$$\hat{\tau} = \tau^{\hat{c}} = \begin{cases} E_0 + E_1 + E_2 + \cdots + E_{m-1} + E_m & \text{if } |x-y| = 2m \\ E_0 + E_1 + E_2 + \cdots + E_{m-1} + E_m + E_{2m+1} & \text{if } |x-y| = 2m+1 \,, \end{cases} \tag{3.1}$$

where $\{E_k\}_{k \geq 0}$ form a set of independent Exponential random variables, with E_k having rate $2k$. (Note that $E_0 \equiv 0$: it is included merely for notational convenience.)

Now define

$$\hat{v}(x, y, t) = \mathbb{P}\left[\hat{\tau} > t \,|\, X_0 = x,\, Y_0 = y\right] \tag{3.2}$$

to be the tail probability of the coupling time under $\hat{c}$. The main result of this paper is the following.

Theorem 3.1. *For any states $x, y \in \mathbb{Z}_2^n$ and time $t \geq 0$,*

$$\hat{v}(x, y, t) = \inf_{c \in \mathcal{C}} \mathbb{P}\left[\tau^c > t \,|\, X_0 = x, Y_0 = y\right] \,. \tag{3.3}$$

In other words, $\hat{\tau}$ is the stochastic minimum of all co-adapted coupling times for the pair (X, Y).

It is clear from the representation in (3.1) that $\hat{v}(x, y, t)$ only depends on (x, y) through $|x - y|$, and so we shall usually simply write

$$\hat{v}(k, t) = \mathbb{P}\left[\hat{\tau} > t \,|\, N_0 = k\right] \,,$$

with the convention that $\hat{v}(k,t) = 0$ for $k \leq 0$. Note, again from (3.1), that $\hat{v}(k,t)$ is strictly increasing in k. For a strategy $c \in \mathcal{C}$, define the process S^c_t by

$$S^c_t = \hat{v}\left(X^c_t, Y^c_t, T-t\right)\,,$$

where $T > 0$ is some fixed time. This is the conditional probability of X and Y not having coupled by time T, when strategy c has been followed over the interval $[0,t]$ and $\hat{c}$ has then been used from time t onwards. The optimality of $\hat{c}$ will follow by Bellman's principle (see, for example, [7]) if it can be shown that $S^c_{t\wedge\tau^c}$ is a submartingale for all $c \in \mathcal{C}$, as demonstrated in the following lemma. (Here and throughout, $s \wedge t = \min\{s,t\}$.)

Lemma 3.1. *Suppose that for each $c \in \mathcal{C}$ and each $T \in \mathbb{R}_+$,*

$$(S^c_{t\wedge\tau^c})_{0\leq t\leq T} \quad \textit{is a submartingale.}$$

Then equation (3.3) *holds.*

Proof. Notice that, with $(X_0, Y_0) = (x,y)$, $S^c_0 = \hat{v}(x,y,T)$ and $S^c_{T\wedge\tau^c} = \mathbf{1}_{[T<\tau^c]}$. If $S^c_{\cdot\wedge\tau^c}$ is a submartingale it follows by the Optional Sampling Theorem that

$$\mathbb{P}\left[\tau^c > T\right] = \mathbb{E}\left[S^c_{T\wedge\tau^c}\right] \geq S^c_0 = \hat{v}(x,y,T) = \mathbb{P}\left[\hat{\tau} > T\right]\,,$$

and hence the infimum in (3.3) is attained by $\hat{c}$.

Now, (point process) stochastic calculus yields:

$$dS^c_t = dZ^c_t + \left(\mathcal{A}^c_t\hat{v} - \frac{\partial\hat{v}}{\partial t}\right)dt\,, \tag{3.4}$$

where Z^c_t is a martingale, and $\mathcal{A}^c_t$ is the "generator" corresponding to the matrix $Q^c(t)$. Since the Poisson processes Λ_{ij} are independent, the probability of two or more jumps occurring in the superimposed process $\bigcup\Lambda_{ij}$ in a time interval of length δ is $O(\delta^2)$. Hence, for any function $f : \mathbb{Z}^n_2 \times \mathbb{Z}^n_2 \times \mathbb{R}^+ \to \mathbb{R}$, $\mathcal{A}^c_t$ satisfies

$$\mathcal{A}^c_t f(x,y,t) = \sum_{i=0}^{n}\sum_{j=0}^{n} q^c_{ij}(t)\Big[f(x+e_i, y+e_j, t) - f(x,y,t)\Big]\,.$$

Setting $f = \hat{v}$ gives:

$$\begin{aligned}\mathcal{A}^c_t \hat{v}(x,y,t) &= \sum_{i=0}^{n}\sum_{j=0}^{n} q^c_{ij}(t)\Big[\hat{v}(x+e_i, y+e_j, t) - \hat{v}(x,y,t)\Big] \\ &= \sum_{i=0}^{n}\sum_{j=0}^{n} q^c_{ij}(t)\Big[\hat{v}(\left|x-y+e_i+e_j\right|, t) - \hat{v}(\left|x-y\right|, t)\Big]\,.\end{aligned}$$

In particular, since $\hat{v}$ is invariant under coordinate permutation, if $N^c_t = |x-y| = k$ then

$$\mathcal{A}^c_t \hat{v}(x,y,t) = \sum_{m=-2}^{2} \lambda^c_t(k,k+m)\Big[\hat{v}(k+m,t) - \hat{v}(k,t)\Big]\,, \tag{3.5}$$

where $\lambda^c_t(k,k+m)$ is the rate (according to $Q^c(t)$) at which N^c_t jumps from k to $k+m$. More explicitly,

$$\lambda^c_t(k,k+2) = \sum_{\substack{i,j\in M_t \\ i\neq j}} q^c_{ij}(t)\,, \qquad \lambda^c_t(k,k+1) = \sum_{i\in M_t} \left(q^c_{i0}(t) + q^c_{0i}(t)\right)\,, \tag{3.6}$$

$$\lambda^c_t(k,k-2) = \sum_{\substack{i,j\in U_t \\ i\neq j}} q^c_{ij}(t)\,, \qquad \lambda^c_t(k,k-1) = \sum_{i\in U_t} \left(q^c_{i0}(t) + q^c_{0i}(t)\right)\,, \tag{3.7}$$

and

$$\lambda^c_t(k,k) = \sum_{i\in U_t, j\in M_t} \left(q^c_{ij}(t) + q^c_{ji}(t)\right) + \sum_{i=1}^{n} q^c_{ii}(t)\,. \tag{3.8}$$

It follows from the definition of Q and equations (3.6) to (3.8) that these terms must satisfy the linear constraints:

$$\begin{aligned}&\lambda^c_t(k,k-2) + \frac{1}{2}\lambda^c_t(k,k-1) \leq k\,, \quad \text{and} \\ &\lambda^c_t(k,k-2) + \frac{1}{2}\lambda^c_t(k,k-1) + \lambda^c_t(k,k) + \frac{1}{2}\lambda^c_t(k,k+1) + \lambda^c_t(k,k+2) = n\,.\end{aligned}$$

Denote by L_n the set of non-negative λ satisfying the constraints

$$\lambda(k,k-2) + \frac{1}{2}\lambda(k,k-1) \leq k\,, \quad \text{and} \tag{3.9}$$

$$\lambda(k,k-2) + \frac{1}{2}\lambda(k,k-1) + \lambda(k,k) + \frac{1}{2}\lambda(k,k+1) + \lambda(k,k+2) = n\,. \tag{3.10}$$

Returning to equation (3.4):

$$dS^c_t = dZ^c_t + \left(\mathcal{A}^c_t \hat{v} - \frac{\partial \hat{v}}{\partial t}\right) dt\,.$$

We wish to show that $S^c_{t\wedge\tau^c}$ is a submartingale for all couplings $c \in \mathcal{C}$. We shall do this by showing that $\mathcal{A}^c_t\hat{v}$ is minimised by setting $c = \hat{c}$. This is sufficient because $S^{\hat{c}}_{t\wedge\hat{\tau}}$ is a martingale (and so $\mathcal{A}^{\hat{c}}_t\hat{v} - \partial\hat{v}/\partial t = 0$). Now, from equation (3.5) we know that

$$\mathcal{A}^c_t\hat{v}(k,t) = \sum_{m=-2}^{2} \lambda^c_t(k,k+m)\Big[\hat{v}(k+m,t) - \hat{v}(k,t)\Big]\,.$$

Thus we seek to show that, for all $k \geq 0$ and for all $t \geq 0$,

$$\max_{\lambda\in L_n} \sum_{m=-2}^{2} \lambda(k,k+m)\Big[\hat{v}(k,t) - \hat{v}(k+m,t)\Big] \geq 0\,. \tag{3.11}$$

For each t, this is a linear function of non-negative terms of the form $\lambda(k,k+m)$. Thanks to the monotonicity in its first argument of $\hat{v}$, the terms appearing in the left-hand-side of (3.11) are non-positive if and only if m is non-negative. Hence we must set

$$\lambda(k,k+1) = \lambda(k,k+2) = 0 \tag{3.12}$$

in order to achieve the maximum in (3.11).

It now suffices to maximise

$$\lambda(k,k-1)\Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] + \lambda(k,k-2)\Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big]\,, \tag{3.13}$$

subject to the constraint in (3.9).

Combining (3.9) and (3.13) yields the final version of our optimisation problem:

$$\text{maximise} \qquad \lambda(k,k-1)\left(\Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] - \frac{1}{2}\Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big]\right) \tag{3.14}$$

$$\text{subject to} \qquad 0 \leq \lambda(k,k-1) \leq 2k\,. \tag{3.15}$$

The solution to this problem is clearly given by:

$$\lambda(k,k-1) = \begin{cases} 2k & \text{if } \Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] > \frac{1}{2}\Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big] \\ 0 & \text{otherwise}\,. \end{cases} \tag{3.16}$$

These observations may be summarised as follows:

Proposition 3.1. *For $\lambda \in L_n$, the maximum value of*

$$\sum_{m=-2}^{2} \lambda(k,k+m)\Big[\hat{v}(k,t) - \hat{v}(k+m,t)\Big]\,,$$

is achieved at λ^*, *where* λ^* *satisfies the following:*

$$\lambda^*(k,k+1) = \lambda^*(k,k+2) = 0\,;$$
$$\lambda^*(k,k-2) + \frac{1}{2}\lambda^*(k,k-1) = k\,;$$
$$\lambda^*(k,k-1) = \begin{cases} 2k & \textit{if } \Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] > \frac{1}{2}\Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big] \\ 0 & \textit{otherwise}\,. \end{cases}$$

Our final proposition shows that $\lambda^*(k,k-1) = 2k$ if and only if k is odd.

Proposition 3.2. *For any fixed* $t \geq 0$,

$$2\Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] - \Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big] \geq 0 \quad \textit{if } k \textit{ is odd, and} \tag{3.17}$$
$$2\Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] - \Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big] \leq 0 \quad \textit{if } k \textit{ is even.} \tag{3.18}$$

Proof. Define $\hat{V}_\alpha$ by

$$\hat{V}_\alpha(k) = \int_0^\infty e^{-\alpha t}\hat{v}(k,t)dt = \frac{1}{\alpha}\left(1 - \mathbb{E}\left[e^{-\alpha\hat{\tau}}\right]\right)\,.$$

We also define $d(k,t) = \hat{v}(k,t) - \hat{v}(k-1,t)$, and for $\alpha \geq 0$ let

$$D_\alpha(k) = \int_0^\infty e^{-\alpha t} d(k,t)dt$$

be the Laplace transform of $d(k,\cdot)$. Given the representation in equation (3.1) of $\hat{\tau}$ as a sum of independent Exponential random variables, it follows that

$$\hat{V}_\alpha(k) = \begin{cases} \dfrac{1}{\alpha}\left(1 - \displaystyle\prod_{i=1}^m \frac{2i}{2i+\alpha}\right) & \text{if } k = 2m \\ \dfrac{1}{\alpha}\left(1 - \dfrac{2(2m+1)}{2(2m+1)+\alpha}\displaystyle\prod_{i=1}^m \frac{2i}{2i+\alpha}\right) & \text{if } k = 2m+1\,. \end{cases} \tag{3.19}$$

To ease notation, let

$$\phi_\alpha(m) = \prod_{i=1}^m \frac{2i}{2i+\alpha}\,.$$

The following equality then follows directly from consideration of the transition rates corresponding to strategy $\hat{c}$:

for all $\alpha \geq 0$ and $m \geq 1$,

$$\begin{aligned}1-\alpha\hat{V}_\alpha(2m)+2m\left[\hat{V}_\alpha(2m-2)-\hat{V}_\alpha(2m)\right] &= \phi_\alpha(m)+\frac{2m}{\alpha}\left[\phi_\alpha(m)-\phi_\alpha(m-1)\right] \\ &= \phi_\alpha(m)+\frac{2m}{\alpha}\phi_\alpha(m)\left[1-\frac{2m+\alpha}{2m}\right] \\ &= 0\,. \end{aligned} \tag{3.20}$$

Similarly,

$$1-\alpha\hat{V}_\alpha(2m-1)+2(2m-1)\left[\hat{V}_\alpha(2m-2)-\hat{V}_\alpha(2m-1)\right]=0\,. \tag{3.21}$$

Now suppose that $k=2m$, and hence is even. We wish to prove that

$$d(2m-1,t)-d(2m,t)\geq 0 \quad \text{for all } t\geq 0\,,$$

which is equivalent to showing that $D_\alpha(2m-1)-D_\alpha(2m)$ is totally (or completely) monotone (by the Bernstein-Widder Theorem; Theorem 1a of [3], Ch. XIII.4).

We proceed by subtracting equation (3.21) from (3.20):

$$\begin{aligned}0 &= -\alpha\left[\hat{V}_\alpha(2m)-\hat{V}_\alpha(2m-1)\right]+2m\left[\hat{V}_\alpha(2m-2)-\hat{V}_\alpha(2m)\right] \\ &\qquad\qquad +2(2m-1)\left[\hat{V}_\alpha(2m-1)-\hat{V}_\alpha(2m-2)\right] \\ &= -\alpha D_\alpha(2m)-2m\left[D_\alpha(2m)+D_\alpha(2m-1)\right]+2(2m-1)D_\alpha(2m-1)\,,\end{aligned}$$

and so

$$D_\alpha(2m-1)-D_\alpha(2m)=\frac{2+\alpha}{2m-2}D_\alpha(2m)\,. \tag{3.22}$$

It therefore suffices to show that $(2+\alpha)D_\alpha(2m)$ is completely monotone.

Now note from the form of $\hat{V}$ in equation (3.19), that

$$(2+\alpha)D_\alpha(2m)=2\Theta_\alpha(2m)\,,$$

where $\Theta_\alpha(2m)$ is the Laplace transform of

$$\theta(2m,t)=\mathbb{P}\left[\sum_{i=0}^{m}E_i>t\right]-\mathbb{P}\left[\sum_{i=0}^{m-1}E_i+E_{2m-1}>t\right]\,,$$

where $\{E_i\}_{i\geq 0}$ form a set of independent Exponential random variables, with E_i having parameter $2i$. But since $\theta(2m,t)$ is strictly positive for all t, it follows that

$(2+\alpha)D_\alpha(2m)$ is completely monotone, as required. This proves that, for any fixed $t \geq 0$,

$$2\Big[\hat{v}(k,t) - \hat{v}(k-1,t)\Big] - \Big[\hat{v}(k,t) - \hat{v}(k-2,t)\Big] \leq 0 \tag{3.23}$$

whenever k is even. Thus inequality (3.18) holds in this case.

Now suppose that $k = 2m+1$, and hence is odd. In this case we wish to show that inequality (3.17) holds, which is equivalent to showing that $D_\alpha(2m+1) - D_\alpha(2m)$ is completely monotone. Now, substituting $m+1$ for m in equation (3.21) yields

$$1 - \alpha\hat{V}_\alpha(2m+1) + 2(2m+1)\left[\hat{V}_\alpha(2m) - \hat{V}_\alpha(2m+1)\right] = 0\,. \tag{3.24}$$

Proceeding as above, we subtract equation (3.20) from (3.24):

$$\begin{aligned} 0 &= -\alpha\left[\hat{V}_\alpha(2m+1) - \hat{V}_\alpha(2m)\right] + 2(2m+1)\left[\hat{V}_\alpha(2m) - \hat{V}_\alpha(2m+1)\right] \\ &\qquad\qquad + 2m\left[\hat{V}_\alpha(2m) - \hat{V}_\alpha(2m-2)\right] \\ &= -\alpha D_\alpha(2m+1) - 2(2m+1)D_\alpha(2m+1) + 2m\left[D_\alpha(2m) + D_\alpha(2m-1)\right]\,. \end{aligned} \tag{3.25}$$

Then it follows from equation (3.22) that

$$(2m-2)D_\alpha(2m-1) = (2m+\alpha)D_\alpha(2m)\,. \tag{3.26}$$

Substitution of equation (3.26) into (3.25) gives

$$0 = (4m+2-\alpha)\left[D_\alpha(2m) - D_\alpha(2m+1)\right] + 2\left[D_\alpha(2m-1) - D_\alpha(2m)\right]\,,$$

and so

$$D_\alpha(2m+1) - D_\alpha(2m) = \frac{2}{4m+2+\alpha}\left[D_\alpha(2m-1) - D_\alpha(2m)\right]\,. \tag{3.27}$$

But, since we have already seen that $D_\alpha(2m-1) - D_\alpha(2m)$ is completely monotone, the right-hand-side of equation (3.27) is the product of two completely monotone functions, and so is itself completely monotone [3], as required.

Now we may complete the

Proof of Theorem 3.1. Thanks to Lemma 3.1 and Proposition 3.1, Proposition 3.2, along with equations (3.12) and (3.16), shows that any optimal choice of $Q(t)$, $Q^*(t)$, is of the following form:

- when N_t is odd:

$$q^*_{i0}(t) = q^*_{0i}(t) = 1 \text{ for all } i \in U_t, \text{ (and so } \lambda^*_t(N_t, N_t - 1) = 2N_t) ,$$
$$q^*_{ii}(t) = 1 \text{ for all } i \in M_t ;$$

- when N_t is even:

$$q^*_{i0}(t) = q^*_{0i}(t) = q^*_{ii}(t) = 0 \text{ for all } i \in U_t, \text{ (and so } \lambda^*_t(N_t, N_t - 1) = 0) , \quad (3.28)$$
$$q^*_{ii}(t) = 1 \text{ for all } i \in M_t .$$

This is in agreement with our candidate strategy $\hat{Q}$ (recall Definition 3.1). From equation (3.28) it follows that the values of $q^*_{ij}(t)$ for distinct $i, j \in U_t$ must satisfy

$$\sum_{\substack{i,j \in U_t \\ i \neq j}} q^*_{ij}(t) = |U_t| \, ,$$

but are not constrained beyond this. Our choice of

$$\hat{q}_{ij}(t) = \frac{1}{|U_t| - 1}$$

satisfies this bound, and so $\hat{c}$ is truly an optimal co-adapted coupling, as claimed.

Remark 3.1. Observe that when $k = 1$, equation (3.1) implies that $\hat{v}(1, t) = \hat{v}(2, t)$ for all t. The optimisation problem in (3.14) and (3.15) simplifies in this case to the following:

$$\text{maximise} \quad \lambda(1, 0)\hat{v}(1, t) \quad (3.29)$$
$$\text{subject to} \quad \frac{1}{2}\lambda(1, 0) + \lambda(1, 1) + \frac{1}{2}\lambda(1, 2) \leq n \, . \quad (3.30)$$

As above, this is achieved by setting $\lambda(1, 0) = 2$. Note from equation (3.30), however, that when $k = 1$ there is no obligation to set $\lambda(1, 2) = 0$ in order to attain the required maximum. Indeed, due to the equality between $\hat{v}(1, t)$ and $\hat{v}(2, t)$, when $k = 1$ it is not sub-optimal to allow *matched* coordinates to evolve independently (corresponding to $\lambda^c_t(1, 2) > 0$), so long as strategy $\hat{c}$ is used once more as soon as $k = 2$.

4. Maximal coupling

Let X and Y be two copies of a Markov chain on a countable space, starting from different states. The coupling inequality (see, for example, [8]) bounds the tail

distribution of *any* coupling of X and Y by the total variation distance between the two processes:

$$\|\mathcal{L}(X_t) - \mathcal{L}(Y_t)\|_{TV} \leq \mathbb{P}\left[\tau > t\right] . \tag{4.1}$$

Griffeath [5] showed that, for discrete-time chains, there always exists a *maximal* coupling of X and Y: that is, one which achieves equality for all $t \geq 0$ in the coupling inequality. This result was extended to general continuous-time stochastic processes with paths in Skorohod space in [11]. However, in general such a coupling is not co-adapted. In light of the results of Section 3, where it was shown that $\hat{c}$ is the optimal co-adapted coupling for the symmetric random walk on $\mathbb{Z}_2^n$, a natural question is whether $\hat{c}$ is also a maximal coupling.

This is certainly not the case in general. Suppose that X and Y are once again random walks on $\mathbb{Z}_2^n$, with $X_0 = (0, 0, \dots, 0)$ and $Y_0 = (1, 1, \dots, 1)$: calculations as in [2] show that the total variation distance between X_t and Y_t exhibits a cutoff phenomenon, with the cutoff taking place at time $T_n = \frac{1}{4}\log n$ for large n. This implies that a maximal coupling of X and Y has expected coupling time of order T_n. However, it follows from the representation of $\hat{\tau}$ in equation (3.1) that

$$\mathbb{E}\left[\,\hat{\tau}\ ;\ |X_0 - Y_0| = n = 2m\right] = \mathbb{E}\left[E_1 + E_2 + \cdots + E_{m-1} + E_m\right] \sim \frac{1}{2}\log(n)\,. \tag{4.2}$$

It follows that $\hat{c}$ is not, in general, a maximal coupling.

A faster coupling of X and Y was proposed by [9]. This coupling also makes new coordinate matches in pairs, but uses information about the future evolution of one of the chains in order to make such matches in a more efficient manner. This coupling is very near to being maximal (it captures the correct cutoff time), but is of course not co-adapted. Further results related to the construction of maximal couplings for general Markov chains may be found in [4, 6, 10].

References

[1] Aldous, D. (1983). *Random walks on finite groups and rapidly mixing Markov chains* vol. 986 of *Lecture Notes in Math.* Springer, Berlin. pp. 243–297.

[2] Diaconis, P., Graham, R. L. and Morrison, J. A. (1990). Asymptotic analysis of a random walk on a hypercube with many dimensions. *Random Structures Algorithms* **1,** 51–72.

[3] Feller, V. (1971). *An introduction to probability theory and its applications* second ed. vol. 2. Wiley.

[4] Greven, A. (1987). Couplings of Markov chains by randomized stopping times. *Probab. Theory Related Fields* **75,** 195–212 and 431–458.

[5] Griffeath, D. (1975). A maximal coupling for Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **31,** 95–106.

[6] Harison, V. and Smirnov, S. N. (1990). Jonction maximale en distribution dans le cas markovien. *Probab. Theory Related Fields* **84,** 491–503.

[7] Krylov, N. V. (1980). *Controlled diffusion processes* vol. 14 of *Applications of Mathematics*. Springer-Verlag, New York. Translated from the Russian by A. B. Aries.

[8] Lindvall, T. (2002). *Lectures on the coupling method*. Dover.

[9] Matthews, P. (1987). Mixing rates for a random walk on the cube. *SIAM J. Algebraic Discrete Methods* **8,** 746–752.

[10] Pitman, J. W. (1976). On coupling of Markov chains. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete* **35,** 315–322.

[11] Sverchkov, M. Y. and Smirnov, S. N. (1990). Maximal coupling of D-valued processes. *Soviet Math. Dokl.* **41,** 352–354.